中国人口增长预测模型

摘要

本文从中国的实际情况和中国人口增长特点出发，参考相关数据，建立了中国人口增长的数学模型，并由此对中国人口增长的中短期和长期趋势做出了分析和预测。

利用 Logistic 曲线预测模型指出了未来几十年内中国人口总量的变化趋势：在未来一段时间内，中国人口仍将处于增长阶段，并在 21 世纪的中叶中国人口总数将达到峰值 $P_m = 15.4707(亿)$ 。

根据过去 10 年来中国人口总量，利用 GM(1,1)模型灰色预测法预测 2006~2010 年短期中国人口总量预测：131890，132803，133722，134647 和 135579 万人。通过残差检验，其平均相对误差仅为 0.11%；对模型进行关联度检验和后验查检验，其结果均优于灰色预测精度检验等级标准。

在实际生活中，人口年龄分布状况在人口预测中是十分重要的，因为不同年龄人的生育率和死亡率是有着很大的差别的，这些对人口发展具有重要作用。对此，根据人口密度函数和不同年龄段的生育率和死亡率建立人口发展方程模型。该模型为人口密度函数的一阶偏微分方程，在数值计算中，将其离散化，即建立离散型方程模型。根据该模型即可预测出任意年代人口密度函数，即可得到该年代的人口指数，能准确地反应出该年代人口的总量和特点。该模型还着重考虑了人口生育率和生育模式对人口发展的影响。当总和生育率 $\beta(t)$ 为 1.8 时，在未来二十多年里，中国人口总量仍然是增长趋势，将在 2027 年左右达到峰值；中国 15~64 岁的劳动年龄人口将在 2013 年将达到高峰 9.83 亿人，但随着老龄人口的增加，劳动年龄人口将会减少；中国 65 岁以上老龄人口总量在 2006 年为 1.2 亿，到 2040 年达到峰值，为 3.06 亿，占总人口的 21.6%，社会将进入老龄化。当 $\beta(t)$ 分别为 1.5，1.8 和 2.0 时，老龄化指数和中国人口总量在 2050 年分别上升到 0.6092，0.5649, 0.5383 和 12.7，13.7，15.1 亿人。为了使得中国人口总量避免过速增长和老龄化指数过高，政府应该将总和生育率控制在 1.8 左右。

最后，建立人口迁移模型，综合考虑出生率、死亡率、生育率、年龄结构、迁移人口等多因素对人口总量及其他指标的影响。模型结果表明，在未来 20 年，中国将有约达 2.6 亿的农村人口陆续转化为城镇人口，大量的乡村人口涌入城市给城市的基础设施及城市管理能力带来了挑战。如果不加控制，未来中国出生人口性别比将一直处于偏高水平，并且有持续升高的势头，出生人口性别比依旧是中国未来人口工作中面临的大问题。各地区生育率均呈下降趋势，老年人口比重日趋上升，目前，中国 60 岁以上老年人口占总人口的 9.43%，到 2020 年，60 岁以上老年人口比重增长到 15.6%，农村老龄化形势更为严峻，政府尤其要关注农村庞大老年人群中的贫困化和边缘化问题。

关键字：Logistic 曲线预测；GM(1,1)灰色预测；人口发展方程；人口迁移

问题重述

中国是一个发展中国家，又是世界上人口最多的国家，人口问题一直是制约中国经济和社会发展的首要因素，因此，能否对中国人口增长做出准确分析和预测，对于加速推进中国现代化建设有着极为重要的现实意义。

近年来中国的人口发展出现了一些新的特点，例如，老龄化进程加速、出生人口性别比持续升高，以及乡村人口城镇化等因素，这些都影响着中国人口的增长。

现有 2001 至 2005 年的市、镇和乡人口不同性别的人在该类人口中所占的百分比，各类年龄段人口的死亡率，以及各类年龄段妇女生育率等数据，参考上述相关数据（也可以搜索相关文献和补充新的数据），从中国的实际情况和人口增长的上述特点出发，建立中国人口增长的数学模型，并由此对中国人口增长的中短期和长期趋势做出预测；特别要指出模型中的优点与不足之处。

问题分析

根据历年来中国人口数量发展特点，发现其增长趋势与 Logistic 曲线相似，故可利用 Logistic 曲线预测未来中国人口长期发展趋势，求出中国人口的峰值。灰色预测模型属于全因素的非线性拟合外推类方法，而人口变化是由很多因素决定的，社会制度、自然环境、生活水平、科学文化水平等都能严重影响社会人口发展过程。这种变化与 GM(1,1)模型具有较强的相似性，故可以运用 GM(1,1)模型的灰色预测法来对中国人口发展趋势做出短期预测。进一步对人口年龄分布进行研究，从人口结构、生育率和死亡率建立人口发展方程，求出任意时刻人口密度函数，即可求得人口的参数指标。进一步考虑实际因素，发现人口的迁移对人口发展也具有重要影响。故建立人口迁移模型，求得人口各个参数指标的发展趋势。

模型建立与求解

1. Logistic 曲线预测模型

1.1 符号说明

t	离起始年 1994 年的年数
$P(t)$	第 t 年中国人口总数
P_0	考察起始年当年的人口总数
P_m	中国人口总量峰值

1.2 模型建立

假定第 $t_0=0$ 年的人口为 $P(0)$，人口增长率为 r，则第 t 年的人口为

$$P(t)=P(0)(1+r)^t \qquad (1.1)$$

经过简单的数学变换，上面递推公式的通用项可以化为指数形式

$$P(t) = P_0 e^{rt} \qquad (1.2)$$

假定变量连续，求导得到齐次方程

$$\frac{dP(t)}{dt} = rP(t) \qquad (1.3)$$

考虑到自然资源、环境条件等因素对人口的增长起阻滞作用，并且随着人口的增加，阻滞作用越大，即人口增长率 r 为人口总量 $P(t)$ 的减函数。因此对上式做出修改，加上一个表征环境约束因子的二次项 $qP(t)^2$，从而得到二阶 Bernoulli 式齐次方程

$$\frac{dP(t)}{dt} = rP(t) - qP(t)^2 = rP(t)[1 - \frac{P(t)}{P_m}] \qquad (1.4)$$

这里 q 为约束参数，$P_m = r/q$ 表示区域饱和人口即最大人口容量。该方程的初始条件和饱和条件分别为 $P(t_0) = P_0, P(t) \le P_m$。解之得到 $Logistic$ 预测模型

$$P(t) = \frac{P_m}{1 + (\frac{P_m}{P_0} - 1)e^{-rt}} = \frac{P_m}{1 + \lambda e^{-rt}} \qquad (1.5)$$

上式中参数 $\lambda = (P_m/P_0 - 1)$。

1.3 模型求解

因为公式(1.5)中只有三个参数(P_m、λ、r)需要求解，所以只需要三个年份的 $P(t)$ 值便可反解出 P_m、λ、r 的值。利用附录 2 中人口数据，随机选取三个年份，由表中给出城市、城镇及乡村中男女人数可确定出当年份的人口总数，例如 1995 年、2004 年、2005 年人口总数分别为 P(1)=12.111、P(10)=12.9988、P(11)=13.075（亿），求出三个参数分别为

$$\begin{cases} P_m = 15.4707 \\ \lambda = 0.265678 \\ r = 0.0371571 \end{cases}$$

将 P_m、λ、r 的值代入公式(1.5)中，则每给出一个年份(t值)，则可对应求出当年的人口总数。我们用 matlab 求解后，预测结果如下：

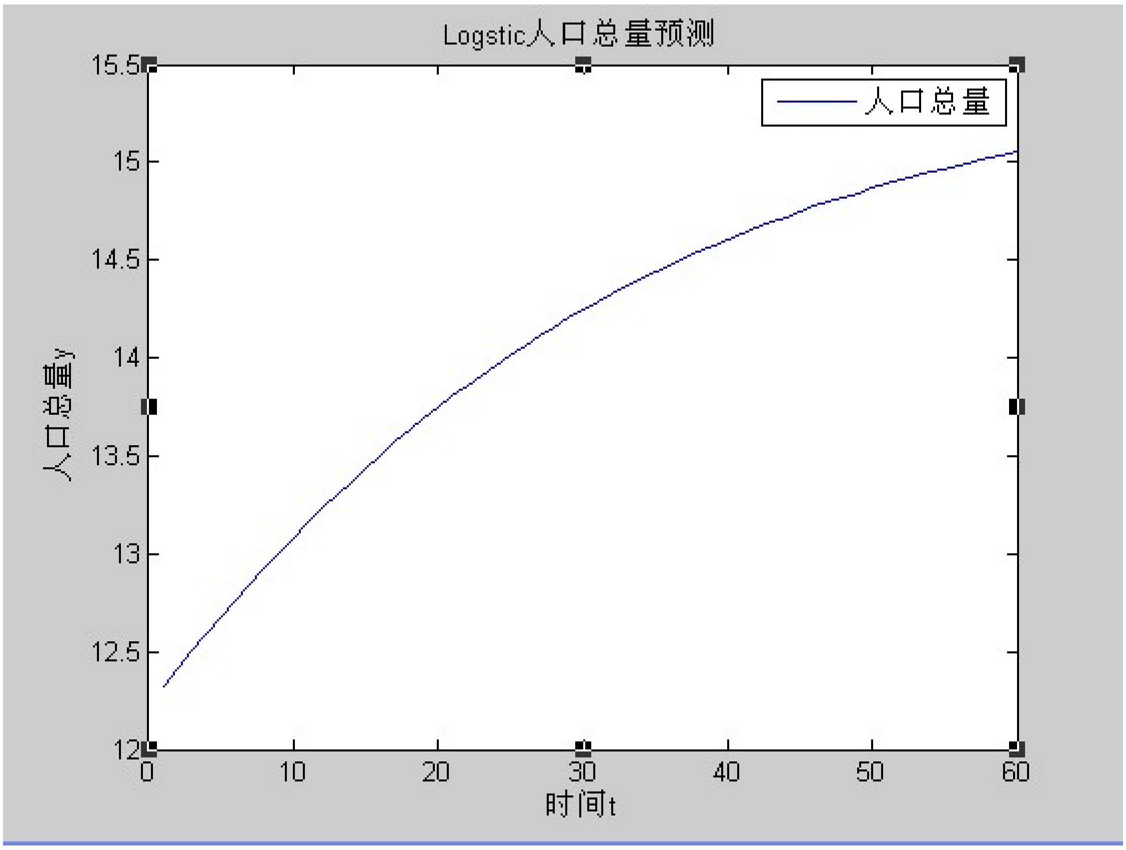

图 1　　96 年到 2055 年的人口变化趋势

由上分析可知，在 21 世纪的中叶中国人口总数将达到峰值 $P_m = 15.4707$(亿)，此后人口总量将趋于稳定，即增长率趋于零。

1.4 模型评价

Logistic 模型简单易行，当有足够历史数据时，其参数能较容易地求出，并且在实际应用中与历史数据吻合地很好。但是，其只考虑人口总数和总的增长率，不涉及人口年龄结构，这将导致其无法预测人口特征一些参数，如年龄结构，男女性别比例等。

2. GM(1,1)模型灰色预测人口总量

一个国家的人口变化是由很多因素决定的，社会制度、自然环境、生活水平、科学文化水平等都能严重影响社会人口发展过程。如此众多的因素不可能通过几个指标就能表达清楚，它们对人口增长的潜在而复杂的影响更是无法精确计算的。因此，我们考虑用灰色预测模型来解决。灰色预测模型属于全因素的非线性拟合外推类方法，在形式上是单数列预测，只运用研究对象自身的时间序列建立模型，与其相关联的因素没有参与建模，这正是灰色系统“灰”的体现。

2.1 GM(1,1)模型建立步骤

$GM(1,1)$ 模型是基于累加生成的数列预测模型，建立模型步骤[3]：

（1）$x^0(1), x^0(2), ..., x^0(M)$ 是所要预测的某项指标的原始数据，对原始数据做一次累加生成处理，得到：

$$x^{(1)}(M) = \sum_{i=1}^{M} x^{(0)}(t), \tag{2.1}$$

得到一个新的数列。这个新的数列与原始数列相比，其随机性程度大大弱化了，而且平稳性大大增加。

（2）将新数列的变化趋势近似用微分方程描述：

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u, \tag{2.2}$$

其中，a, u 为待定参数，利用最小二乘法拟合得到：

$$\begin{bmatrix} a \\ u \end{bmatrix} = (B^T B)^{-1} B^T Y_{M.} \tag{2.3}$$

(3) 构造数据矩阵。方程（2.3）中 $Y_{M.}$ 为列向量，$Y_M = [x^0(2), x^0(3), ..., x^0(M)]^T$，$B$ 为构造数据矩阵，

$$B = \begin{bmatrix} -\frac{1}{2}[x^1(1) + x^1(2)] & 1 \\ -\frac{1}{2}[x^1(2) + x^1(3)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[x^1(M-1) + x^1(M)] & 1 \end{bmatrix} \tag{2.4}$$

（4）求出预测模型：

$$x^{(1)}(t+1) = [x^{(0)}(1) - \frac{u}{a}]e^{-at} + \frac{u}{a} \tag{2.5}$$

2.2 灰色预测模型建立

人口增长主要有自然增长和机械增长两种方式，利用 1996~2005 年中国全国人口的统计资料(表 1)[2]，对未来 5 年内全国人口做出预测。

表 1　　1996~2005 年全国人口总量（万人）

年份	1996	1997	1998	1999	2000
人口总量	122389	123626	124761	125786	126743
年份	2001	2002	2003	2004	2005
人口总量	127627	128453	129227	129988	130756

由表一可得全国人口总量的原始时间序列为： $x^0(t)$={122389,123626,124761,125786,126743,127627,128453,129227,129988,130756}

由方程(1)对原始数据做一次累加生成处理，得到：$x^{(1)}(t)$={122389，246015，370776，496562，623305，750932，879385，1008612，1138600，1269356}；

由方程（4）得：

$$B=\begin{bmatrix} -184202 & 1 \\ -308395.5 & 1 \\ -433669 & 1 \\ -559933.5 & 1 \\ -687118.5 & 1 \\ -815158.5 & 1 \\ -943998.5 & 1 \\ -1073606 & 1 \\ -1203978 & 1 \end{bmatrix}$$

Y_M ={123626,124761,125786,126743,127627,128453,129227,129988,130756}；

由方程（2.3）得：

$$\begin{bmatrix} a \\ u \end{bmatrix}=\begin{bmatrix} -0.006896 \\ 122682.2792 \end{bmatrix}$$

由方程（2.5）可得人口总量预测模型：

$$x^{(1)}(t+1)=(1.797e^{0.006896t}-1.779)\times 10^7 \qquad (2.6)$$

2.3 模型检验

2.3.1 残差检验

将 $t=0,1,...,9$ 代入预测方程（2.6）中，得到 1996~2005 年全国人口总量的一次累加值，

$q(t)$：第 t 年预测值与实际值的相对误差；

$\hat{\varepsilon}^{(0)}(t)$：第 t 年预测值与实际值的绝对误差；

$\hat{x}^{(0)}(t)$：第 t 年原始预测值；

$\hat{x}^{(1)}(t)$：第 t 年第一次累加预测值。

即由

$$\hat{x}^{(0)}(t)=\hat{x}^{(1)}(t)-\hat{x}^{(1)}(t-1)$$
$$\hat{\varepsilon}^{(0)}(t)=x^{(0)}(t)-\hat{x}^{(0)}(t)$$
$$q(t)=\frac{\hat{\varepsilon}^{(0)}(t)}{\hat{x}^{(0)}(t)}\times 100\%$$

分别求出预测值，绝对误差值和相对误差值，见表 2：

表 2　全国人口总量实际值与预测值（万人）绝对误差和相对误差

年份	1996	1997	1998	1999	2000
实际值	122389	123626	12461	125786	126743
预测值	122389	123953.2	124811	125674.7	126544.4
绝对误差	0.0	327.2	50.0	111.3	198.6
相对误差	0.00	0.26	0.04	0.09	0.16
年份	2001	2002	2003	2004	2005
实际值	127627	128453	129227	129988	130756
预测值	127420.1	128301.9	129189.7	130083.8	130983.9
绝对误差	206.9	151.1	37.3	95.8	227.9
相对误差	0.16	0.12	0.03	0.07	0.17

由表 2 可知，平均相对误差为 0.11%，模型精度比较高。

2.3.2　关联度检验

令关联系数 $\xi_i=\frac{\Delta_{\min}+k\Delta_{\max}}{\Delta_i+\Delta_{\max}}(t=1,2,...,10)$, 其中 $\Delta_{\min}=\min\Delta_i,\Delta_{\max}=\max\Delta_i$,

$\Delta_i=\left|x_i^{(1)}(t)-x_i^{(0)}(t)\right|$。从表 2 数据得：

$\min\{\Delta_i\}$={0，327.2 ，50.0 ，111.3 ，198.， 206.9 ，151.1 ，37.3 ，95.8，227.9}，

$\max\{\Delta_i\}$={0，327.2 ，50.0 ，111.3 ，198.， 206.9 ，151.1 ，37.3 ，95.8，227.9}，

由此计算灰数为 0.5 的关联数：

ξ_i={1，0.5101，0.8675，0.7462，0.6223，0.6127，0.6841，0.8978，0.7736，0.5894}。

关联度 $\gamma=\frac{1}{n}\sum_{i=1}^{n}\xi(i)$, 则 $x^{(0)}(t)$与$x^{(1)}(t)$的关联度为 $\gamma=0.73$。

满足 $\rho=0.5$时，$\gamma>0.6$的检验标准。

2.3.3　后验查检验

由原始数据序列 $x^{(0)}(t)$ 和绝对误差序列 $\Delta(t)$ 计算得原始数据序列和绝对误差序列的

标准差分别为：

$$S_1=\sqrt{\frac{\sum\left[x^{(0)}(t)-\bar{x}^{(0)}\right]^2}{n-1}}=2782 \qquad S_2=\sqrt{\frac{\sum\left[\Delta^{(0)}(t)-\bar{\Delta}^{(0)}\right]^2}{n-1}}=101.03$$

由此计算方差比 $C=\frac{S_2}{S_1}=0.0363$，小误差概率 $P=\{\left|\Delta^{(0)}(t)-\bar{\Delta}^t\right|<0.6745S_1\}$. 设 $M=\left|\Delta^{(0)}(t)-\bar{\Delta}^t\right|$，则 $M=\{140.61，186.62，90.61，29.32，57.99，66.28，10.52，103.35，44.8594，87.3387\}$。设 $S=0.6745S_1$,则$S=1870$，所有 M 均小于 S ,s 所以 $P=1$。

表 3　灰色预测精度检验等级标准

检验指标	好	合格	勉强	不合格
P	>0.95	>0.80	>0.70	<0.70
C	<0.35	<0.50	<0.65	>0.65

将检验指标 P 和 C 与灰色预测精度检验等级标准（表 3）对比可知，预测模型较理想。因此，上述模型可以用于预测。将 t=11,12,13,14,15 代入方程（2.6），经过计算得 2006~2010 年全国人口总量预测值，见表 4。

表 4　2006~2010 全国人口总量预测值　　单位：万人

年份	2006	2007	2008	2009	2010
人口	131890	132803	133722	134647	135579

2.4 模型评价

优点：由上述分析可知，GM(1,1)预测模型的数据量要求小，精度高，具有较强的实用性和有效性，是个比较好的预测方法，对于开放性，非线性的复杂系统，GM(1,1)预测模型能够从整体出发对外延不确定性系统变化进行动态的科学模拟与仿真。

局限性：灰色预测模型的可靠性及预测精度主要取决于原始数据列的光滑性，原始数据列的光滑性越好灰色模型的预测精度越高，如果原始数列的光滑性不够，那灰色预测就不很精确，效果不理想。

3. 人口系统预测

前面讨论的 Logistic 模型和灰色预测模型只考虑人口总数和总的增长率，不涉及年龄结构。而事实上，在人口预测中人口年龄分布状况是十分重要的，因为不同年龄人的生育率和死亡率是有着很大的差别的。两个国家或地区目前的人口数相同，但由于年龄结构的差异，那么两者的人口发展状况将大不一样，所以本模型将考虑人口年龄分布对人口预测的影响。

3.1 人口发展方程模型[1]

人口数量的年龄结构变化的因素主要由出生、死亡、迁移。分析全国人口数量和年龄结构时，迁移人口是相对进出境人口而言的，相对于中国庞大的人口基数来说其影响

不大，因此略去不计。

为了研究任意时刻不同年龄的人口数量，引入人口的分布函数和密度函数。时刻t年龄小于r的人口称为人口分布函数，记作$F(r,t)$，其中，$t,r(r\geq 0)$均为连续变量，设F是连续，可微的。时刻t的人口总量记作$N(t)$，最高年龄记作r_m，理论推导时$r_m\to\infty$，于是对于非负非降函数$F(r,t)$有：

$$F(0,t)=0, F(r_m,t)=N(t)$$

人口密度函数定义为：

$$p(r,t)=\frac{\partial F}{\partial r}$$

$p(r,t)dr$表示时刻t年龄在区间$[r,r+dr)$内的人数。

记$\mu(r,t)$为时刻t年龄r的人的死亡率，其含义为：$\mu(r,t)p(r,t)dr$表示时刻t年龄在区间$[r,r+dr)$内单位时间死亡人数。

考虑时刻t年龄在 $[r,r+dr)$内 人到达时刻$t+dt$的情况，他们中活着的那一部分人的年龄变为$[r+dr,r+dr+dr_1)$，这里$dr_1=dt$，而在dt在这段时间内死亡人数为$\mu(r,t)p(r,t)drdt$，于是有

$$p(r,t)dr-p(r+dr_{1,}t+dt)=\mu(r,t)p(r,t)drdt$$

上面公式可以写作：

$$[p(r+dr_1,t+dt)-p(r,t+dt)]+[p(r,t+dt)-p(r,t)]=-\mu(r,t)p(r,t)dt,\qquad dt=dr_1$$

即可得到人口密度函数$p(r,t)$的一阶偏微分方程：

$$\frac{\partial p}{\partial r}+\frac{\partial p}{\partial t}=-\mu(r,t)p(r,t)$$

上述方程有两个定解的条件：初始密度函数记作$p(r,0)=p_0(r)$；单位时间出生的婴儿数记作$p(0,t)=f(t)$，称为婴儿出生率。所以可以将上述方程以及定解条件写作：

$$\begin{cases}\dfrac{\partial p}{\partial r}+\dfrac{\partial p}{\partial t}=-\mu(r,t)p(r,t)\\ p(r,0)=p_0(r),\ r\geq 0\\ p(0,t)=f(t),\quad t\geq 0\end{cases}\qquad (*)$$

这个连续型人口发展方程描述了人口的演变过程，从这个方程确定了密度函数 $p(r,t)$ 后，立即可得出各个年龄的人口数，即人口分布函数

$$F(r,t)=\int_0^r p(s,t)ds$$

在社会安定的局面下和不太长的时间内，死亡率大致与时间无关，于是近似假设 $\mu(r,t)=\mu(r)$，这时方程（*）的解为：

$$p(r,t)=\begin{cases} p_0(r-t)e^{-\int_{r-t}^{r}\mu(s)ds}, 0\le t\le r \\ f(t-r)e^{-\int_0^r \mu(s)ds}, \quad t>r \end{cases}$$

在数值计算时，需要对方程（*）离散化，得到下面的离散模型[5]：

$$\begin{cases} x(t+1)=H(t)x(t)+\beta(t)B(t)x(t) \\ x(0)=[x_1(0),x_2(0),...,x_m(0)]^T \end{cases} \qquad t\ge 0$$

其中，$x(t)=[x_1(t),x_2(t),...,x_{r_m}(t)]^T$，$x_i(t)$ 为年龄 i 在 t 年代的人口数，m 为实际计算的最大年龄，$i=1,2,...,m$，$\beta(t)$ 为在年代 t 的总和生育率；

$$H(t)=\begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1-u_1(t) & 0 & \cdots & 0 & 0 \\ 0 & 1-u_2(t) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1-u_{m-1}(t) & 0 \end{bmatrix} \qquad B(t)=\begin{bmatrix} 0 & b_{r_1}(t) & \cdots & b_{r_2}(t) & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

其中

$$b_i(t)=(1-u_0(t))(1-u_i(t))k_i(t)h_i(t), i=r_1,...,r_2, \qquad (3.1)$$

$u_i(t)$ 为年龄 i 在年代 t 的死亡率，$k_i(t)$ 为年龄 i 在年代 t 中女性所占的比例，$h_i(t)$ 为年龄 i 的妇女在年代 t 中的生育模式，其中 $i=r_1,...,r_2$，$u_0(t)$ 为出生婴儿在年代 t 的死亡率。

3.2 人口发展方程中参数函数的确定

对模型进行数值求解之前，需要确定改方程中的参数函数，即确定死亡率函数 $u_i(t)$，总和生育率 $\beta(t)$，女性性别比例函数 $k_i(t)$ 以及生育模式 $h_i(t)$。

3.2.1 生育模式函数的确定

从方程（3.1）中可以看出，生育模式 $h_i(t)$ 为年龄 i 的妇女在年代 t 时生育加权因子。

在社会和生育政策稳定的环境下，一般认为生育模式 $h_i(t)$ 近似与年代 t 无关，即 $h_i(t)=h_i$， h_i 表示了哪些年龄生育率高，哪些年龄生育率低。

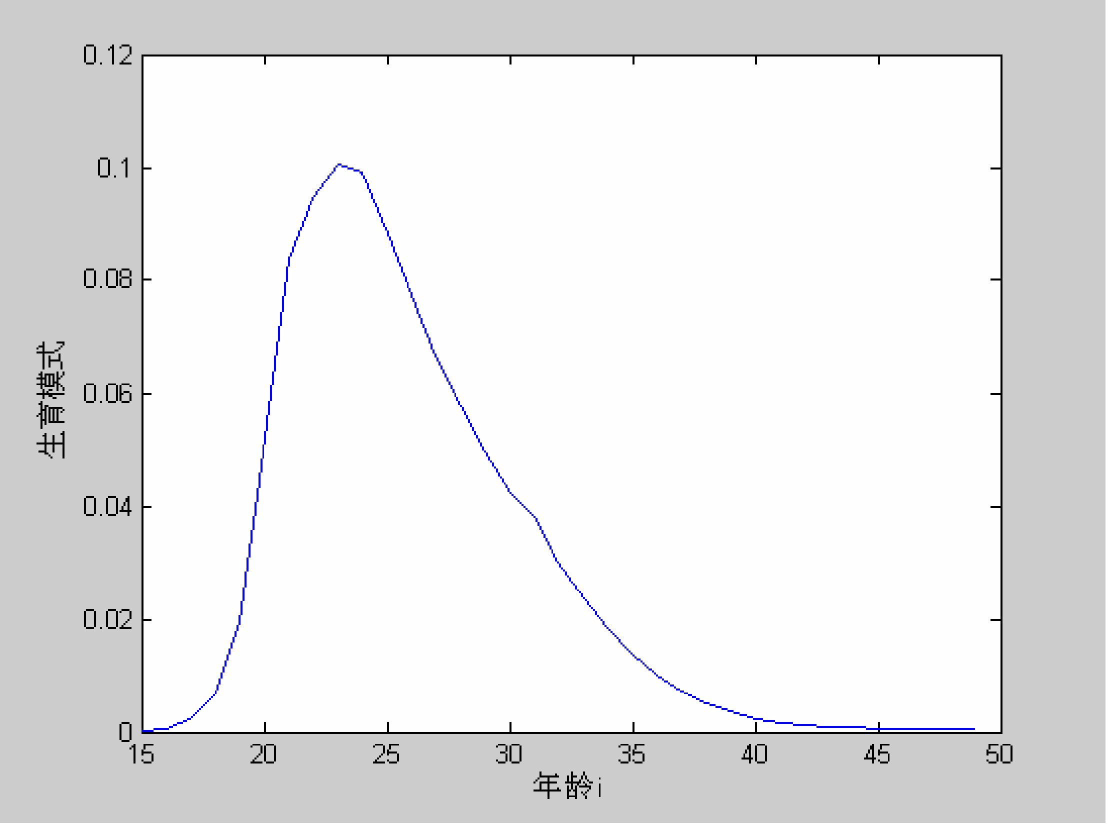

图 2　　生育模式 h_i 示意图

由附件数据（2005 年全国市，镇，乡妇女生育率）综合得到 2005 年 1 全国 5~49 岁女性生育模式 h_i 示意图，育龄区间为[15,49],生育高峰大约在 23 岁左右。做理论分析时，对于生育模式，可以比较准确的用 χ^2 概率密度函数来刻画[5]，因此有:

$$h_i=\begin{cases}\dfrac{0.8(i-r_1)^{\frac{n}{2}-1}e^{\frac{i-r_1}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} & i>r_1\\ 0 & i<r_1\end{cases}$$

根据附件数据，取开始生育年龄 $r_1=15$ 岁， $r_2=49$ 岁；

由 χ^2 概率密度函数的性质得 $r_c=r_1+n-2$，其中 r_c 为生育高峰，由图 1 知， $r_c=23$，得到 $n=11$。

3.2.2 女性性别比例和死亡率函数的确定

对于女性性别比例，由于人们生育观念和政策法律的不断完善，人为因素对女性性别比例的干扰日益减少，女性性别比例主要由自然因素影响，即可以近似认为在未来一段时间内，中国人口女性性别比例保持不变，我们根据 2005 年各个年龄女性所占总人口比例作为未来一段时间内中国各个年龄段女性所占总人口比例。

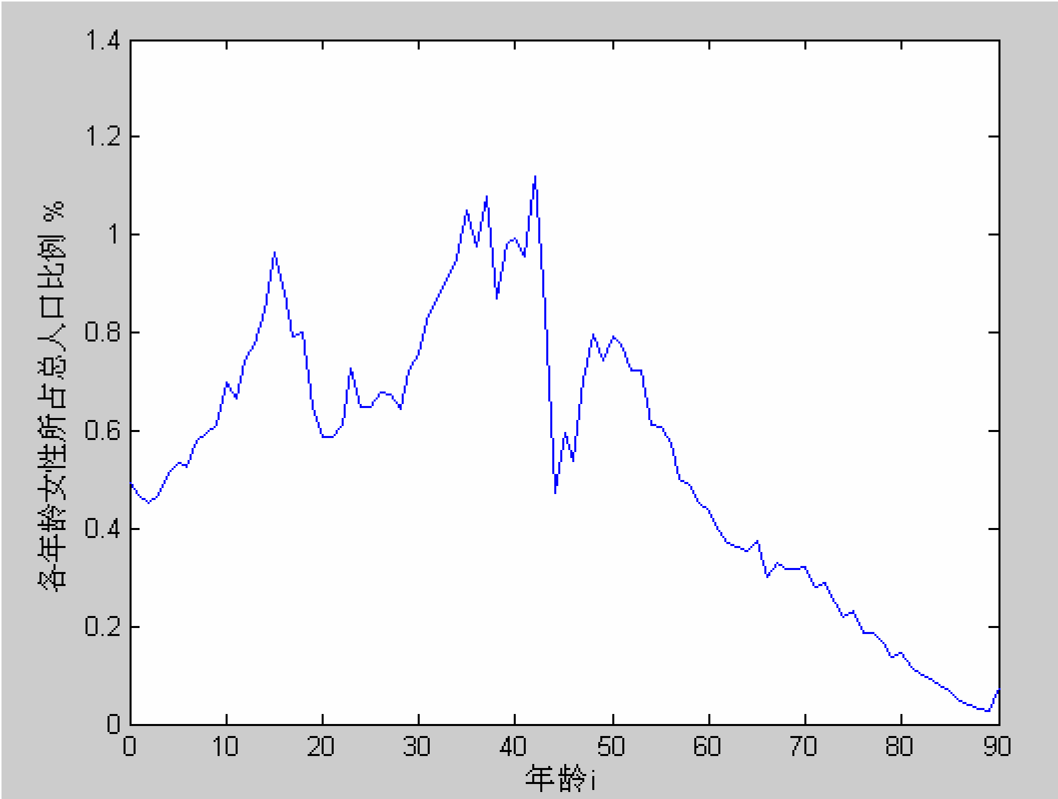

图 3　　2005 年各个年龄女性所占总人口比例(%)

死亡率函数是人口发展方程中非常关键的参数函数，其对模型有较大的影响。从下表可以看出，自1978年以来，中国人口死亡率波动不大，考虑到在社会稳定，科技发展，医疗条件的改善和人民生活水平的提高，未来中国人口死亡率有可能略有下降，但是下降速度并不明显。所以，认为未来一段时间内，中国人口死亡率大约在6.0‰~6.5‰之间，取其平均得到未来一段时间内中国人口死亡率大约为6.25‰，死亡率大致与年代无关，有 $u_i(t)=u_i$。

认为在未来一段时间内，中国人口各个年龄死亡率分布变化不大，即可以按照 2005 年的数据近似计算。

表5　1978~2005年各年全国人口死亡率[2]：（‰）

年代	死亡率	年代	死亡率
1978	6.25	1993	6.64
1980	6.34	1994	6.49
1981	6.36	1995	6.57
1982	6.6	1996	6.56
1983	6.9	1997	6.51
1984	6.82	1998	6.5
1985	6.78	1999	6.46
1986	6.86	2000	6.45
1987	6.72	2001	6.43
1988	6.64	2002	6.41
1989	6.54	2003	6.4
1990	6.67	2004	6.42
1991	6.7	2005	6.51
1992	6.64		

由附件数据（2005 年（市，镇，乡）各个年龄段男性，女性死亡率分布）综合计算得 2005 年全国人口各个年龄段死亡率密度，如图：

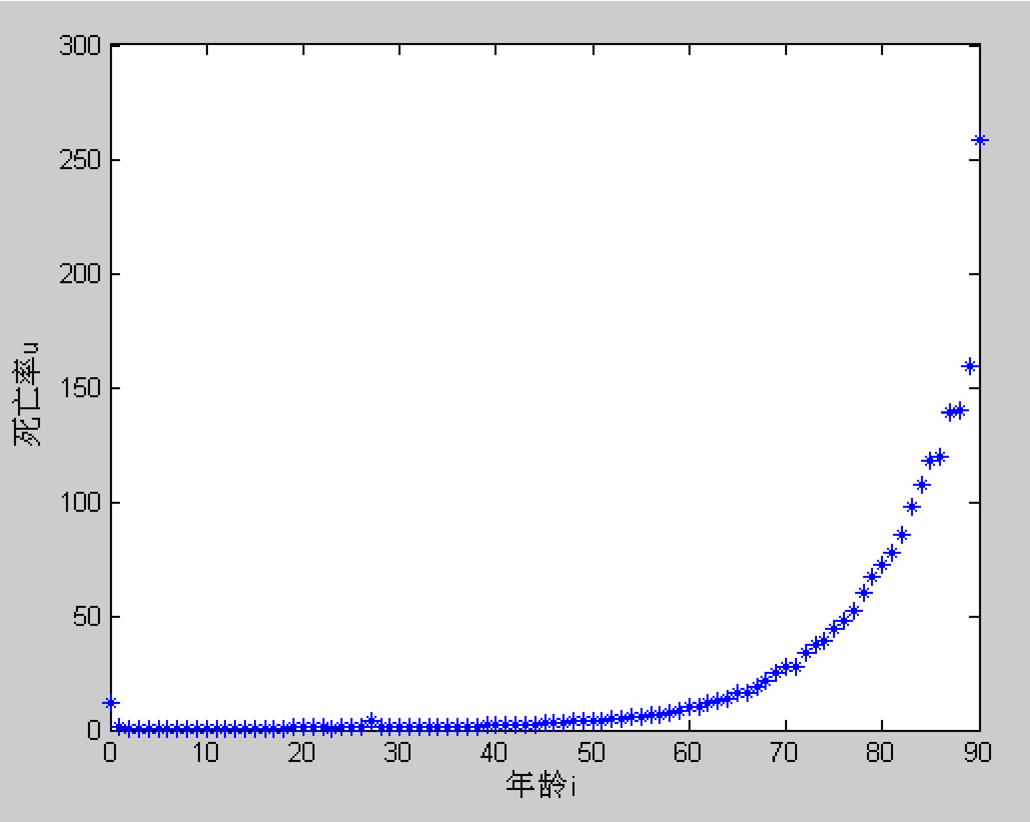

图 4　　2005 年全国人口各个年龄段死亡率密度（‰）

认为在未来一段时间内，中国人口各个年龄死亡率分布变化不大，即可以按照 2005 年的数据近似计算。

3.2.3 总和生育率的确定

如果我们把人口发展方程看作一个动力系统，对方程（*），从控制论的观点来看人口系统，如图：[1]

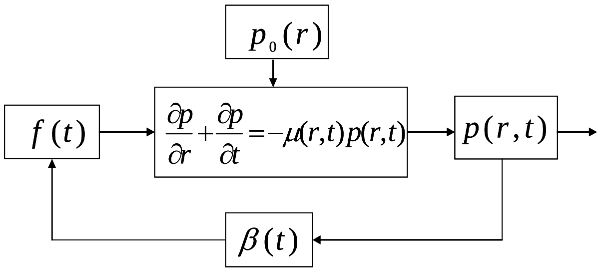

图 5　　人口正反馈控制系统示意图

总和生育率 $\beta(t)$ 是该系统的控制量。由于 $\beta(t)$ 是可控量， $\beta(t)$ 是反馈增益，而且这种反馈还具有相当大的滞后作用。不同的取值将对应不同的预测结果。建立了人口发展过程预测模型之后，人口预测实质就是先给定控制量 $\beta(t)$，然后定量算出人口状态及相应的人口随时间t 变化的趋势。在实际计算中，我们对 $\beta(t)$ 分别取不同的值来得到不同的预测结果，讨论 $\beta(t)$ 对人口的影响。

人口指数[1]：

1. 年代 t 人口总数 $N(t)$：

连续型表达式：$N(t)=\int_0^{r_m}p(r,t)dr$，离散化表达式为 $N(t)=\sum_{i=0}^{m}x_i(t)$

2. 年代 t 平均年龄 $R(t)$：

连续型表达式：$R(t)=\frac{1}{N(t)}\int_0^{r_m}rp(r,t)dr$，离散化表达式为 $R(t)=\frac{1}{N(t)}\sum_{i=0}^{m}ix_i(t)$

3. 年代 t 平均寿命 $S(t)$；

连续型表达式：$S(t)=\int_t^{\infty}e^{-\int_0^{\tau-t}\mu(r,t)dr}d\tau$，离散化表达式为 $S(t)=\sum_{i=0}^{m}e^{-\sum_{k=0}^{i}u_k(t)}$

4. 年代 t 老龄化指数 $\omega(t)$，定义为：

$$\omega(t)=\frac{R(t)}{S(t)}$$

平均年龄 $R(t)$ 越大，$\omega(t)$ 也越大；对于 $R(t)$ 相同的两个国家或地区，平均寿命 $S(t)$ 越大，表示该国家或者地区健康水平越高，也说明一个人在工作的时间在一生中所占比例也越大，于是老龄化指数 $\omega(t)$ 较小。

3.3 模型预测结果

根据人口发展方程模型，由附录 1 得知在 20 世纪 90 年代中后期，中国总和生育率已降到 1.8 左右，并稳定至今. 按照生育率为 1.8，对 2005~2050 年全国人口总量，15~64 岁劳动年龄人口总量和 65 岁以上老龄人口总量分别预测，得到下图：

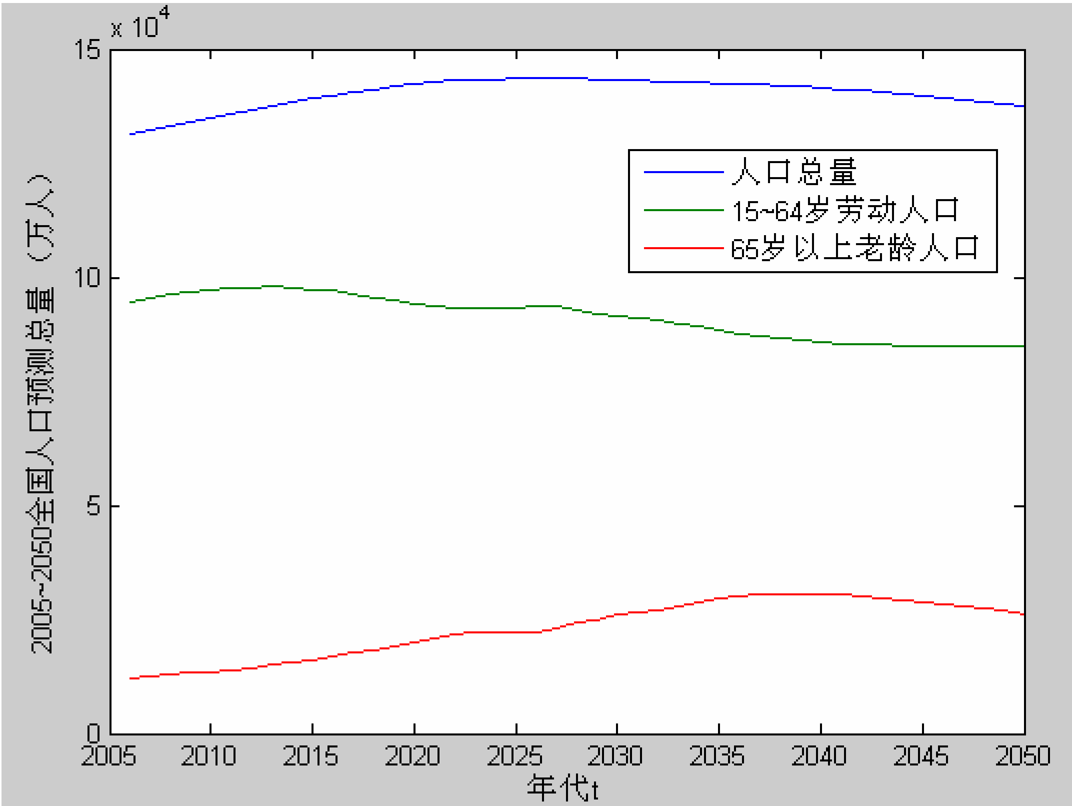

图 6　　2005~2050 年 2005~2050 年全国人口总量、15~64 岁劳动年龄人口总量和 65 岁以上老龄人口总量 （万人）

从图中可以看出，在未来二十多年里，中国人口仍然是增长趋势，将在 2027 年左右达到峰值；中国 15~64 岁的劳动年龄人口将在 2013 年将达到高峰 9.83 亿人，但随着老龄人口的增加，劳动年龄人口将会减少；中国 65 岁以上老龄人口总量在 2006 年为 1.2 亿，到 2040 年达到峰值，为 3.06 亿，占总人口的 21.6%，届时每 3~4 人中就有 1 名老年人，社会将进入老龄化。人口老龄化将导致抚养比不断提高，对社会保障体系和公共服务体系的压力加大。

3.4 模型灵敏度分析

在这我们主要讨论总和生育率 $\beta(t)$ 对人口数量的影响；

把人口发展方程看作一个动力系统，则总和生育率 $\beta(t)$ 是该系统的控制量。由于 $\beta(t)$ 是可控量，不同的取值将对应不同的预测结果。我们对 $\beta(t)$ 分别取不同的值来得到不同的预测结果，如何控制 $\beta(t)$，使得中国人口朝着健康，稳定的方向发展。

对 $\beta(t)=1.5,1.8,2.0$，分别预测到 2050 年中国人口总量和中国老龄化指数，结果如下：

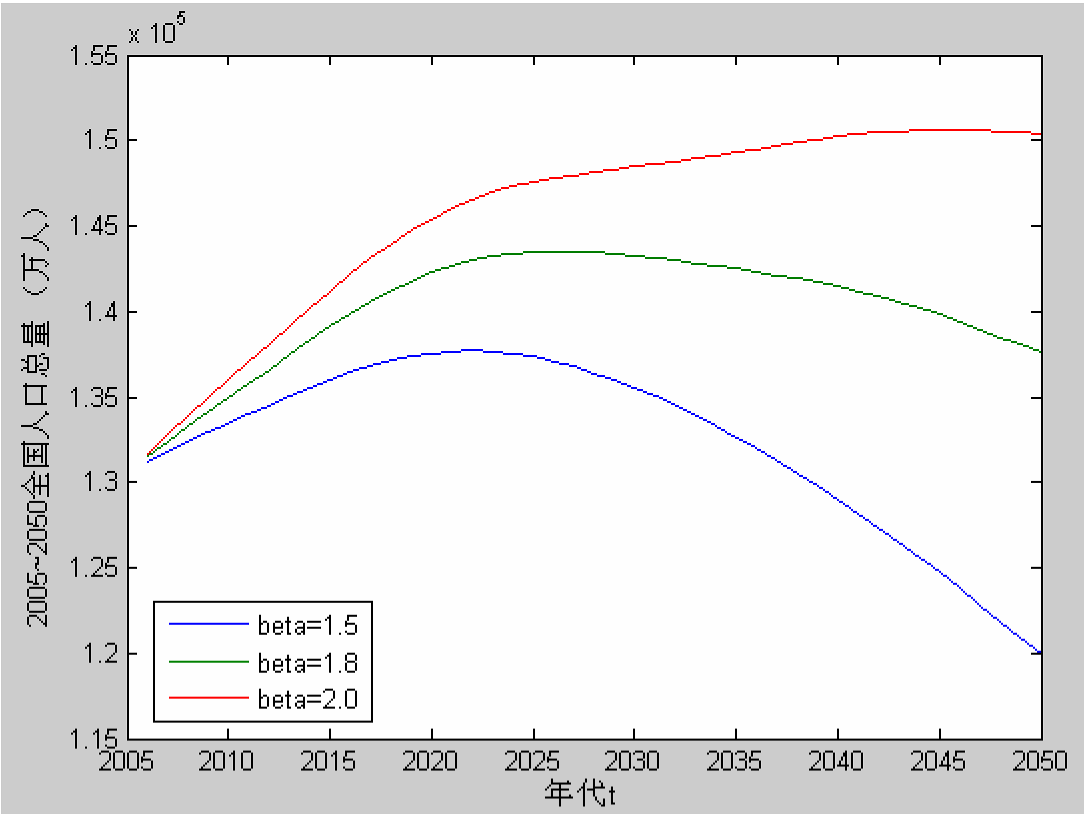

图 7　　2005~2050 年全国人口总量　（万人）

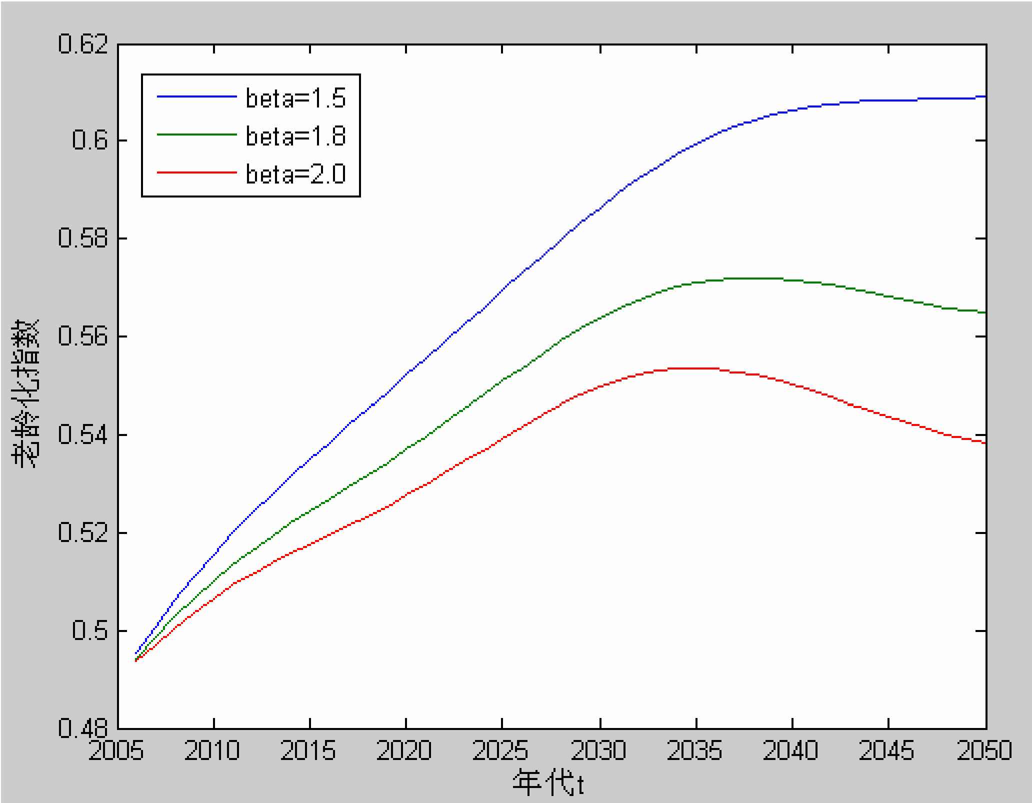

图 8　　2005~2050 年全国人口老龄化指数

从图7，中国人口将在2030年之前仍处于增长期，当β取2.0时，在2050年前，中国人口仍然会缓慢增加。这表明中国人口形势仍然严峻，若取消计划生育政策，人口将会增长更快。庞大的人口总量，对中国现在已经不堪重负的各种资源，以及就业形势将是个极其严峻的考验。所以，计划生育政策在未来二三十年内仍应该坚持。

随着总和生育率β取值的不同，$\beta = 1.5, 1.8, 2$,老龄化指数在2050年分别上升到0.6092，0.5649, 0.5383；按照国际通用的老龄化社会标准，中国已经开始进入老龄化社会阶段且发展速度很快，而且随着时间的推移，老龄化程度将越来越严重, 社会老龄化问题必须得到足够的重视。

3.5 模型评价

人口发展方程模型着重从人口的年龄结构和不同年龄的生育率和死亡率等系统内部因素来预测未来人口的发展趋势。根据人口的内部变化机理，建立起一阶偏微分方程，并通过数值计算，将模型离散化，最终求出结果。所得结果与实际相符，具有良好的预测能力。由于缺少一些实际数据，为了求解的方便快捷，对一些实际问题做了简单的假设，对模型的结果具有一定的影响。比如:随着生活水平的提高和人民生育观念的改变，生育模式和生育率都会随着时间有所改变，育龄妇女的人数所占总人口比例也会有所改变。希望能获得更多的实际数据，在模型迭代过程中不断修正误差，相信将会得到更加理想的预测结果。

4. 人口迁移模型

在 Logistic 曲线预测模型及 GM(1,1)灰色预测模型中，我们只考虑人口总数和总的增长率，不涉及年龄结构。在人口发展方程模型中我们增加考虑了人口年龄分布对人口预测的影响。事实上，不仅年龄结构会影响人口的预测，人口迁移对中国的人口预测也有重大影响。中国是一个发展中国家，城镇化是中国近年来的必然趋势，每年都会有大量的乡村人口涌入城镇，进而影响城镇人口的固有结构。城镇的生活水平及医疗水平高于乡村，导致城镇人口死亡率低于乡村人口死亡率，这使得迁移人口对城镇固有人口结构的影响传递到其对中国人口总数量的影响，并进一步传递到对人口各项指标的影响。本模型将综合考虑出生率、死亡率、生育率、年龄结构、迁移人口等多因素对人口总量及其他指标的影响。

4.1 符号说明

t	年份（$t=0$表示 2005 年）
$P(t)$	第t年的总人口
$P(x,t)$	第t年x岁人口总数
下标r	取"c"表示市，取"t"表示镇，取"v"表示乡，取"ct"表示城镇
下标s	取"f"表示女性，取"m"表示男性
$P_{r,s}(x,t)$	某区域（市镇或乡）第t年x岁男性或女性的人口数
$P_{sum}(t)$	表示 90+岁人口总数
$\rho_{r,s}(x,t)$	某区域中第t年x岁男性或女性占该区域同年龄男性或女性的比例
$d_{r,s}(x,t)$	某区域第t年x岁男性或女性的死亡率
$K_{r,s}(t)$	某区域新生男婴或女婴占当年出生人口的比例
$B_r(t)$	某区域第t年育龄妇女生育率
$b_r(x,t)$	某区域第t年x岁生育子女的妇女占该区域同年龄妇女数比例
$\gamma(t)$	第t年城镇人口占总人口的比例

4.2 模型假设

（1） 各年份的人口统计时间为年底，并且人口迁移均集中在此时。
（2） 每名妇女每次生育最多只生一个小孩。
（3） 2005 年以后的年份各区域各年龄层的死亡率均不变，皆与 2005 年的相应数据相等。

（4） 把数据中年龄为 90+的人口数按 90 岁人口统计，即除死亡人口外 90 岁的老人排年龄不再增长，仍为 90。

（5） 当迁移人口进入某区域后，进入该区域的迁移人口按照年龄进入该地区的年龄分布。该地区死亡率和生育率保持不变。

4.3 模型建立

使人口数量和结构变化的因素主要有出生、死亡和迁移。根据影响人口变动的各要素关系，我们建立了分年龄、分性别、分区域预测的人口方程。

4.3.1 预处理

为了便于考察乡村人口城镇化，我们将市级区域人口与镇级区域人口并在一起，称为城镇人口。城镇人口的各项指标为市及镇人口指标的加权平均值。即

$$\rho_{ct,s}(x,t)=\rho_{c,s}(x,t)\times\frac{P_{c,s}(t)}{P_{c,s}(t)+P_{t,s}(t)}+\rho_{t,s}(x,t)\times\frac{P_{t,s}(t)}{P_{c,s}(t)+P_{t,s}(t)}$$

$$d_{ct,s}(x,t)=d_{c,s}(x,t)\times\frac{P_{c,s}(t)}{P_{c,s}(t)+P_{t,s}(t)}+d_{t,s}(x,t)\times\frac{P_{t,s}(t)}{P_{c,s}(t)+P_{t,s}(t)}$$

$$K_{ct,s}(t)=K_{c,s}(t)\times\frac{P_{c}(t)}{P_{c}(t)+P_{t}(t)}+K_{t,s}(t)\times\frac{P_{t}(t)}{P_{c}(t)+P_{t}(t)}$$

$$B_{ct}(t)=B_{c}(t)\times\frac{P_{c}(t)}{P_{c}(t)+P_{t}(t)}+B_{t}(t)\times\frac{P_{t}(t)}{P_{c}(t)+P_{t}(t)}$$

$$b_{ct}(x,t)=b_{c}(x,t)\times\frac{P_{c,f}(t)}{P_{c,f}(t)+P_{t,f}(t)}+b_{t}(x,t)\times\frac{P_{t,f}(t)}{P_{c,f}(t)+P_{t,f}(t)}$$

4.3.2 确定第 $t+1$ 年中国总人口

由于总人口等于各年龄男性及女性人口的总和，因此有以下公式

$$P(t+1)=\sum_{x=0}^{88}P(x+1,t+1)+P(0,t+1)+P_{sum}(t+1)$$

$$=\sum_{x=0}^{88}P_{f}(x+1,t+1)+\sum_{x=0}^{88}P_{m}(x+1,t+1)+P_{f}(0,t+1)+P_{m}(0,t+1)+P_{sum}(t+1) \quad (4.1)$$

a.由模型假设可知，人口年龄最大值为上式中 $\sum_{x=0}^{88}P_{f}(x+1,t+1)$ 可为在第 $t+1$ 年时中国女性（除 0 岁女婴）总人口。计算时分为两部分，一部分为第 t 年城镇女性在第 $t+1$ 年的存活数量 $P'_{ct,f}(x+1,t+1)$，另一部分为第 t 年乡村女性在第 $t+1$ 年的存活数量

$P_{v,f}^{'}(x+1,t+1)$。同理可计算出$\sum_{x=0}^{88} P_m(x+1,t+1)$。它们的计算公式如下$(0 \le x \le 88)$：

$$P_s(x+1,t+1) = P_{ct,s}^{'}(x+1,t+1) + P_{v,s}^{'}(x+1,t+1) \tag{4.2}$$

$$P_{ct,s}^{'}(x+1,t+1) = P(t)\times\gamma(t)\times\rho_{ct,s}(x,t)\times[1-d_{ct,s}(x,t)] \tag{4.2.a}$$

$$P_{v,s}^{'}(x+1,t+1) = P(t)\times[1-\gamma(t)]\times\rho_{v,s}(x,t)\times[1-d_{v,s}(x,t)] \tag{4.2.b}$$

b.同样，$P_f(0,t+1)$及$P_m(0,t+1)$的确定也一样，分区域进行考虑。

$$P_s(0,t+1) = P_{ct,s}^{'}(0,t+1) + P_{v,s}^{'}(0,t+1) \tag{4.3}$$

由模型假设中第三条可知，$d_{r,s}(x,t) = d_{r,s}(x,0) \quad (t \ge 0)$

因此第$t+1$年中各区域新生男性或女性婴儿的存活数量$P_{ct,s}^{'}(0,t+1)$可用下式求出

$$P_{r,s}^{'}(0,t+1) = K_{r,s}(t+1)\times[1-d_{r,s}(0,0)]\times\sum_{x=15}^{49} P_{r,f}(x,t)\times b_r(x,t) \tag{4.3.a}$$

c.由于我们假设年龄超过 90 的按 90 计，因此我们用$P_{sum}(t+1)$来对 90+岁的人口进行累加。$P_{sum}(t+1)$的计算方法很简单，即是把前一年 89 岁到当年存活下来的人数加上前一年 90+岁到当年存活下来的人数：

$$P_{sum(r,s)}(t+1) = P_{sum(ct,s)}^{'}(t+1) + P_{sum(v,s)}^{'}(t+1) \tag{4.4}$$

$$P_{sum(ct,s)}^{'}(t+1) = P(t)\times\gamma(t)\times\sum_{x=89}^{90}\left\{\rho_{ct,s}(x,t)\times[1-d_{ct,s}(x,t)]\right\} \tag{4.4.a}$$

$$P_{sum(v,s)}^{'}(t+1) = P(t)\times[1-\gamma(t)]\times\sum_{x=89}^{90}\left\{\rho_{v,s}(x,t)\times[1-d_{v,s}(x,t)]\right\} \tag{4.4.b}$$

由以上系列公式$(4.1)\sim(4.4)$则可确定出中国总人口数量。

4.3.3 确定中国第$t+1$年城镇人口与乡村人口的年龄以及性别结构

先考虑从乡村到城镇的净迁入人口的年龄以及性别结构。

中国是发展中国家，乡村人口城镇化是一种必然存在的现象。为简化问题，我们只考虑人口的净迁移量。每年城镇人口净迁入量为当年的城镇人口减去城镇的固有人口，即第$t+1$年中乡村至城镇人口的净迁移量：

$$\Delta P_{ct}(t+1) = P(t+1)\times\gamma(t) - [P_{ct}(0,t+1) + \sum_{x=0}^{89} P_{ct}(x+1,t+1)] \tag{4.5}$$

由于迁移人口多为青壮年劳动力，依据历史数据，迁移劳动力性别约衡定为 1.6:1;[4] 记性别比为η，表示男性或女性迁移人口占总迁移人口的比例，因此表示男性比例时，η取 61.538%，表示女性比例时，η取 38.462%。假设每年迁移人口的年龄结构服从正态分布 $N(26,8)$ [4];即 $\Delta P_{ct}(x,t)$ 对 x 的分布概率为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-u)^2}{2\sigma^2}}=\frac{1}{\sqrt{16\pi}}e^{-\frac{(x-26)^2}{16}} \tag{4.6}$$

大量的人口迁移带来了城镇及乡村人口性别结构及年龄结构变化。以城镇第$t+1$年底$x+1$岁女性数量为例，该数量分为两部分：一部分为第t年城市x岁女性在当年的存活数量，另一部分为第$t+1$年乡村净迁入到城镇$x+1$岁女性数量；城镇总人口可直接由中国总人口及城镇化比例求出。乡村则相反，求第$t+1$年底$x+1$岁女性数量时，需要把第t年城市x岁女性在当年的存活数量减去乡村第$t+1$年净迁入到城镇的$x+1$岁女性数量。因此，可分别求出第$t+1$年城镇或乡村中$x+1$岁男性或女性人数：

$$P_{ct,s}(x+1,t+1)=P_{ct,s}^{'}(x+1,t+1)+\Delta P_{ct,f}(t+1)\times\eta\times\int_{x+1}^{x+2}f(y)dy \quad (x\geq 0) \tag{4.7.1}$$

$$P_{v,s}(x+1,t+1)=P_{v,s}^{'}(x+1,t+1)-\Delta P_{ct,s}(t+1)\times\eta\times\int_{x+1}^{x+2}f(y)dy \quad (x\geq 0) \tag{4.7.2}$$

$$P_{ct,s}(0,t+1)=P_{ct,s}^{'}(0,t+1)+\Delta P_{ct,f}(t+1)\times\eta\times\int_{0}^{1}f(y)dy \tag{4.7.3}$$

$$P_{v,s}(0,t+1)=P_{v,s}^{'}(0,t+1)-\Delta P_{ct,f}(t+1)\times\eta\times\int_{0}^{1}f(y)dy \tag{4.7.4}$$

从而可求出第$t+1$年城镇总人口及第$t+1$年乡村总人口为

$$P_{ct}(t+1)=P_{ct,f}(0,t+1)+P_{ct,m}(0,t+1)+\sum_{x=0}^{89}P_{ct,f}(x+1,t+1)+\sum_{x=0}^{89}P_{ct,m}(x+1,t+1)$$

$$P_{v}(t+1)=P_{v,f}(0,t+1)+P_{v,m}(0,t+1)+\sum_{x=0}^{89}P_{v,f}(x+1,t+1)+\sum_{x=0}^{89}P_{v,m}(x+1,t+1)$$

4.4 人口差分预测参数设定

1. 某区域新生男婴或女婴占当年出生人口的比例 $K_{r,s}(t+1)$ 确定

   $K_{r,s}(t)$ 主要反映的是出生人口性别比，利用所给的 2001 年至 2005 年的出生人口性别比数据，拟合得到未来 50 年的性别比。

2. 某区域各年龄段育龄妇女的生育分布 $b_r(x,t)$ 的确定

   $b_r(x,t)$ 主要反映了第t年时某区域各年龄段育龄妇女的生育分布情况，利用所给的 2001 年至 2005 年的数据，拟合未来 50 年的某区域各年龄段育龄妇女的生育分布情况。

3. 每年城镇化比例$\gamma(t)$的确定

分析所给数据，得到 2001-2005 年城镇人口占总人口的比例(%)分别为$\{37.66, 39.09, 40.53, 41.76, 42.99\}$,发现城镇化速度每年约递增$\Delta = 1\%$ [4]；

所以可得

$\gamma(t) = \gamma(0) + \Delta \times t$　　　　　$\gamma(0)$为初始城镇化比例

4. $\rho_{r,s}(x,t)$ 的确定

当$x \geq 0$时，有

$$\rho_{ct,s}(x+1,t+1) = \frac{P'_{ct,s}(x+1,t+1) + \Delta P_{ct,f}(t+1) \times \eta \times \int_{x+1}^{x+2} f(y)dy}{P(t+1) \times \gamma(t+1)} \tag{4.8.1}$$

$$\rho_{v,s}(x+1,t+1) = \frac{P'_{v,s}(x+1,t+1) - \Delta P_{ct,s}(t+1) \times \eta \times \int_{x+1}^{x+2} f(y)dy}{P(t+1) \times [1-\gamma(t+1)]} \tag{4.8.2}$$

另外，0 岁的婴儿上诉参数需要做小小改动。即把第一部分的数量改为第$t+1$年中城镇新生婴儿的存活数量$P'_{ct,s}(0,t+1)$。即

$$\rho_{ct,s}(0,t+1) = \frac{P'_{ct,s}(0,t+1) + \Delta P_{ct,f}(t+1) \times \eta \times \int_{0}^{1} f(y)dy}{P(t+1) \times \gamma(t+1)} \tag{4.8.3}$$

$$\rho_{v,s}(0,t+1) = \frac{P'_{v,s}(0,t+1) - \Delta P_{ct,f}(t+1) \times \eta \times \int_{0}^{1} f(y)dy}{P(t+1) \times [1-\gamma(t+1)]} \tag{4.8.4}$$

4.5 模型预测结果

4.5.1　未来中国人口总量预测

未来 10 年人口数量，如下：

表 6　　2006 年~2015 年中国人口总量预测值（亿）

年代	2006	2007	2008	2009	2010
人口总数	13.1106	13.1465	13.1827	13.2192	13.2559
年代	2011	2012	2013	2014	2015
人口总数	13.293	13.3303	13.368	13.4059	13.4441

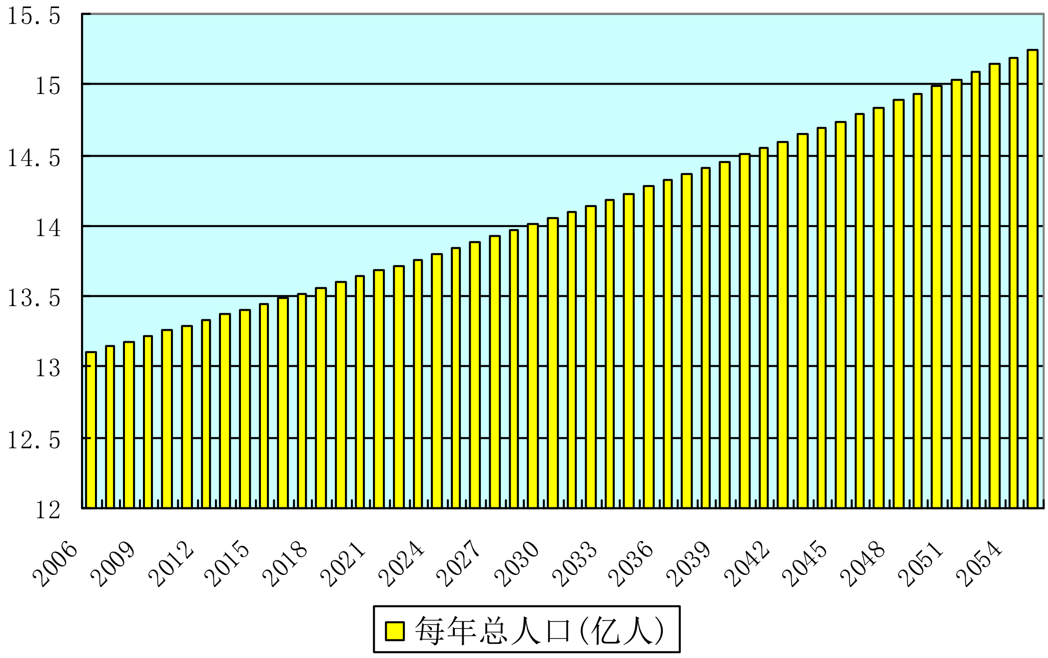

图 9　　未来 50 年中国人口总数变化趋势预测

仿真结果显示，在未来的 50 年，中国人口仍将持续攀升。预计在 21 世纪中叶，中国人口将达到 15 亿左右。

4.5.2　净迁移人口预测

由公式(4.5)可编程计算得每年的净迁移人口数 $\Delta P_{ct}(t)$：

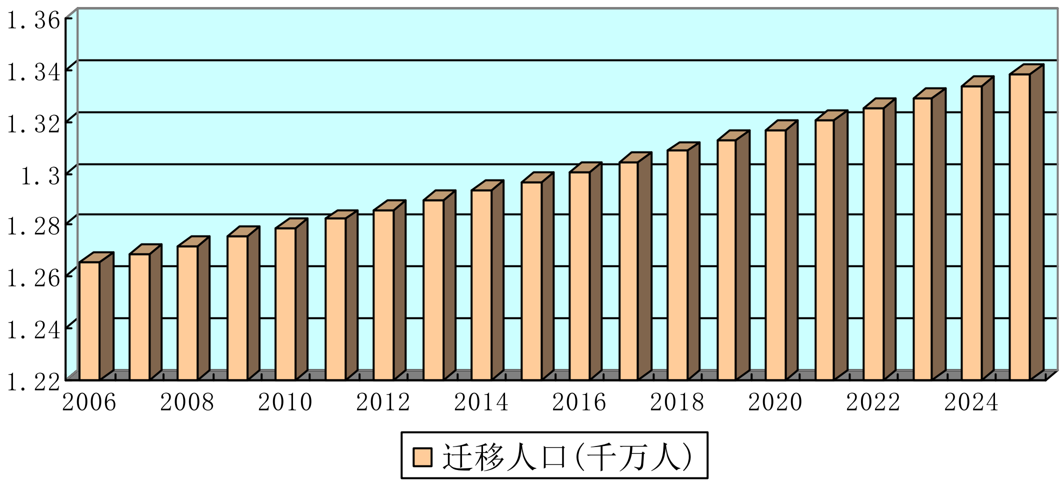

图 10　中国未来 20 年净迁移人口数量预测

在未来 20 年，中国将有约达 2.6 亿的农村人口陆续转化为城镇人口。日益庞大的流入人口对城市的基础设施、公共服务和城市管理能力提出挑战，也增加了计划生育管理

和服务的难度。

4.5.3 出生人口性别比预测

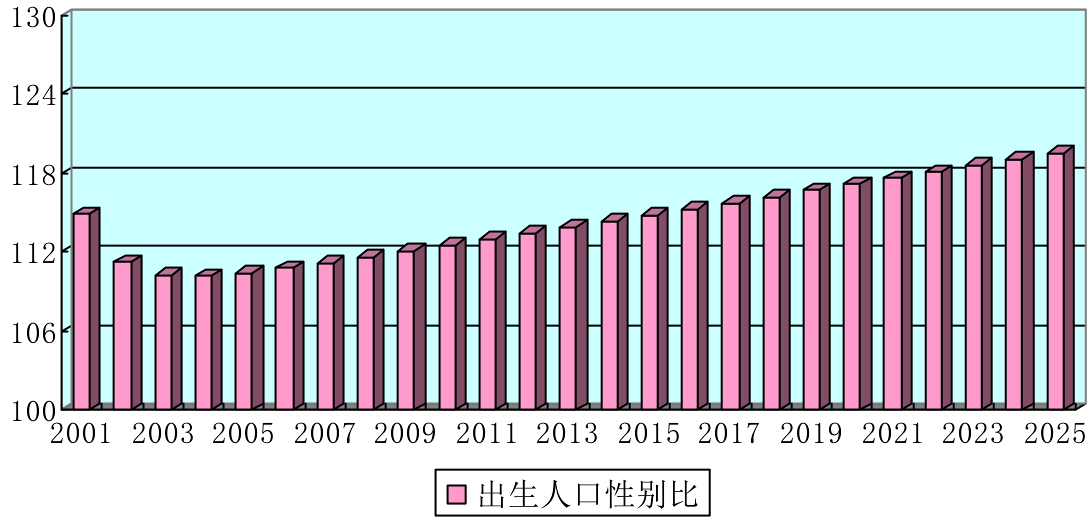

图 11 2001~2025 年中国出生人口性别比预测

从上图明显可以看出，如果不加控制，未来中国出生人口性别比将一直处于偏高水平，并且有持续升高的势头，这表明出生人口性别比依旧是中国未来人口工作中面临的大问题。在未来出生人口性别比持续升高，表明以后新进入婚育年龄人口男性将明显多于女性，婚姻挤压问题凸现，低收入及低素质者结婚难，所导致的社会秩序混乱将成为影响社会稳定的严重隐患。

4.5.4 年龄结构老龄化预测

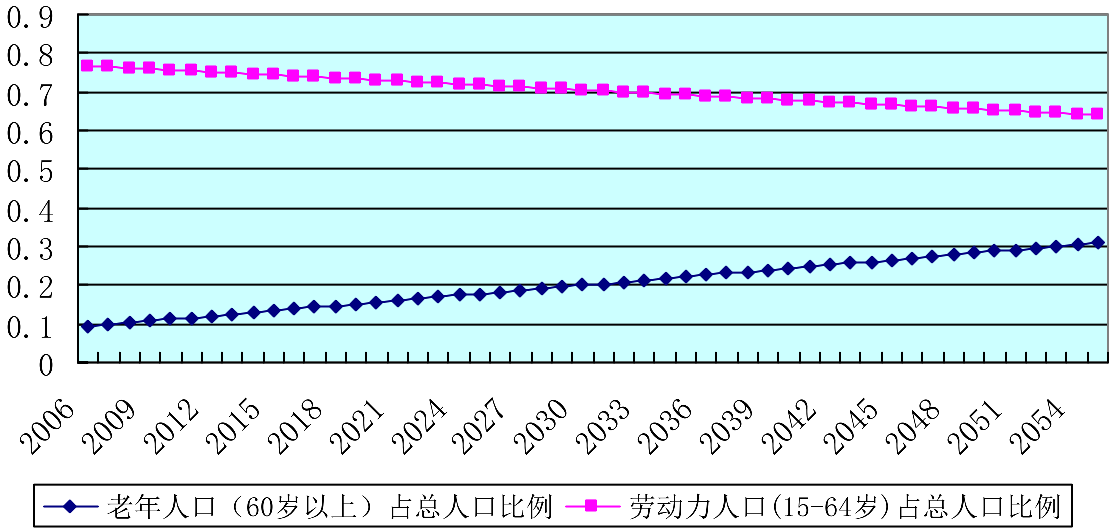

图 12 未来 50 年中国老年人口及劳动力人口比例变化趋势

仿真结果显示，老年人口比重日趋上升，而劳动力人口比例日趋下降。促使人口老龄化的主要原因是生育率降低。农村社会养老保障制度不健全，青壮年人口大量流入城市，使农村老龄化形势更为严峻。政府尤其要关注庞大老年人群中的贫困化和边缘化问题。目前，中国 60 岁以上老年人口已达 1.24 亿，占总人口的 9.43%。到 2020 年，60 岁以上老年人口将达到 2.12 亿人，比重从 2006 年的 9.43%增长到 15.6%。预计本世纪 40 年代后期形成老龄人口高峰平台，60 岁以上老年人口达 4.3 亿人，比重达 30%。届时每 3－4 人中就有 1 名老年人。

4.5.5 生育率变化趋势预测

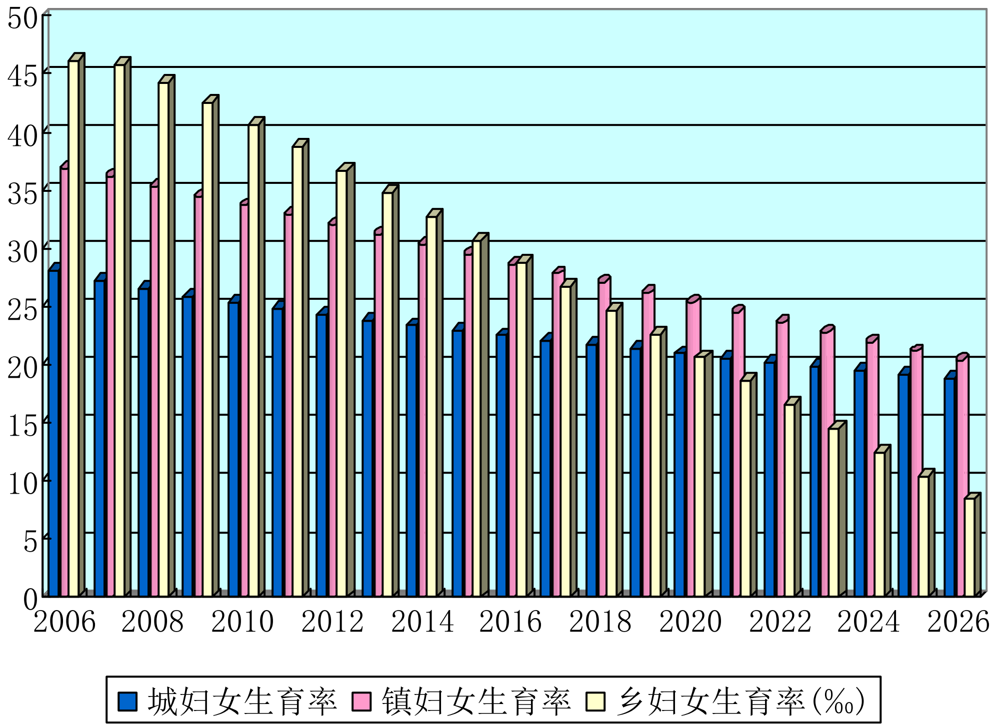

图 13　未来 20 年中国各区域妇女生育率变化趋势

由图 13 可知，生育率总体呈下降趋势。且明显乡村妇女的生育率下降趋势要快于城镇的，这是由迁移人口的年龄构成偏青壮年所造成。

4.6 模型评价

优点：在考虑人口年龄结构，出生率和死亡率等内部因素时，进一步考虑人口迁徙对人口发展的影响。利用差分方程，动态分析人口变化趋势，展现了人口迁徙对人口发展的影响，揭示了大量农村青壮年涌入城镇，对城镇和农村的人口结构，生育率和死亡率的影响。这些对政府加强管理，科学决策具有重要作用。

局限性：由于缺少必要的数据的，我们假设迁移人口男女比固定，未随时间变，未分年龄层；这将导致模型预测精度受到影响。

参考文献

［1］姜启源，谢金星，叶俊 数学模型（第三版）［M］北京：高等教育出版社 2003 年。
［2］中华人民共和国国家统计局 中国统计年鉴［M］北京：中国统计出版社。
［3］周瑞平 $GM(1,1)$ 模型灰色预测法预测城市人口规模［J］内蒙古师范大学学报（自然科学汉文版) 第34卷第一期 2003年5月。
［4］新华社，中国城镇化率已达 43.9%，http://www.cpirc.org.cn/news/rkxw_gn_detail.asp?id=8014，2007 年 03 月 01 日。
［5］屠小明，冯元珍等，基于发展方程的人口系统预测和分析[J]，南京人口管理干部学院学报，第 22 卷第 3 期：25－26 页，2006 年 7 月。